

Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments

Romain Koszul¹, Sandrine Caburet^{2,3},
Bernard Dujon¹ and Gilles Fischer^{1,*}

¹Unité de Génétique Moléculaire des Levures, Département de Structure et Dynamique des Génomes, Institut Pasteur, Paris, France and ²Unité de Stabilité des Génomes, Département de Structure et Dynamique des Génomes, Institut Pasteur, Paris, France

There is growing evidence that duplications have played a major role in eucaryotic genome evolution. Sequencing data revealed the presence of large duplicated regions in the genomes of many eucaryotic organisms, and comparative studies have suggested that duplication of large DNA segments has been a continuing process during evolution. However, little experimental data have been produced regarding this issue. Using a gene dosage assay for growth recovery in *Saccharomyces cerevisiae*, we demonstrate that a majority of the revertant strains (58%) resulted from the spontaneous duplication of large DNA segments, either intra- or interchromosomally, ranging from 41 to 655 kb in size. These events result in the concomitant duplication of dozens of genes and in some cases in the formation of chimeric open reading frames at the junction of the duplicated blocks. The types of sequences at the break-points as well as their superposition with the replication map suggest that spontaneous large segmental duplications result from replication accidents. Aneuploidization events or suppressor mutations that do not involve large-scale rearrangements accounted for the rest of the reversion events (in 26 and 16% of the strains, respectively).

The EMBO Journal (2004) 23, 234–243. doi:10.1038/sj.emboj.7600024; Published online 18 December 2003

Subject Categories: genome stability and dynamics; genomic and computational biology

Keywords: gene dosage recovery; replication; segmental duplication; yeast

Introduction

The change from one gene copy per genome to two copies is a significant mutational event representing the first step leading to the formation of gene families. It is regarded as the

most effective mechanism by which new functions can arise and eventually contribute to the phenotypic differences between species (Prince and Pickett, 2002). Sequence-based evidence for the presence of large segmental duplications has been found in the genomes of many eucaryotic organisms (Wolfe and Shields, 1997; Bailey *et al.*, 2002; Simillion *et al.*, 2002; Thomas *et al.*, 2003), and comparative studies have suggested that duplication of large DNA segments has been a continuing process during evolution (Llorente *et al.*, 2000b; Samonte and Eichler, 2002; Locke *et al.*, 2003). However, it is actively debated whether these large duplicated regions were produced through the accumulation of segmental duplications or by complete genome duplication events (Llorente *et al.*, 2000a,b; Friedman and Hughes, 2001; Robinson-Rechavi *et al.*, 2001; Wong *et al.*, 2002; Blanc *et al.*, 2003; Friedman and Hughes, 2003).

Earlier attempts to characterize gene duplication were mainly based on the yeast *CUP1* gene dosage selection system, and different mechanisms of primary gene amplification, such as aneuploidization, extrachromosomal amplification or cycle of breakage–fusion–bridge, were described (Whittaker *et al.*, 1988; Spector and Fogel, 1992; Dorsey *et al.*, 1993; Moore *et al.*, 2000). Some DNA content variations of large chromosomal regions were reported as potential large segmental duplications in the *Saccharomyces cerevisiae* genome (Bach *et al.*, 1995; Roelants *et al.*, 1995; Hughes *et al.*, 2000). In addition, early work on filamentous fungi genetics revealed large duplications for chromosome segments resulting from nonreciprocal terminal translocations (Bainbridge and Roper, 1966; Sexton and Roper, 1984; Daud *et al.*, 1985). Nonreciprocal translocations have also been described in evolving populations of diploid yeast cells (Dunham *et al.*, 2002).

If segmental duplications have played a significant role in eucaryotic genome evolution, it is unclear how these duplications are generated, intra- or interchromosomally, in direct or inverted orientation, and what sort of sequences are involved at the junctions. Thus, the availability of an experimental system adequately designed to investigate the molecular mechanisms involved in the formation of duplications would be of general relevance for the study of genome evolution. Here, we designed a gene dosage assay in *S. cerevisiae* to monitor gene duplication. The growth defect caused by deletion of one duplicate from a highly similar gene pair can be compensated either by the duplication of the remaining gene of the pair (segmental duplications or aneuploidization) or by other types of suppressor mutations that do not involve large-scale rearrangements. In this work, we focus on the cases of segmental duplications and demonstrate that, while we select for the duplication of a single gene, duplications of large DNA segments encompassing numerous genes occur spontaneously in the *S. cerevisiae* genome. Repair of replication-dependent

*Corresponding author. Unité de Génétique Moléculaire des Levures (CNRS URA2171, UFR927 Université Pierre et Marie Curie), Département de Structure et Dynamique des Génomes, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris cedex 15, France.
Tel.: +33 1 4568 8507; Fax: +33 1 4061 3456;
E-mail: fischer@pasteur.fr

³Present address: INSERM U361/E0021 Hôpital Cochin, Pavillon Baudelocque, 123 bd de Port-Royal, 75014 Paris, France

Received: 5 August 2003; accepted: 3 November 2003; Published online: 18 December 2003

chromosome breakages could be at the origin of these segmental duplications.

Results

A gene dosage recovery system to screen for spontaneous duplication events

The system designed to monitor gene duplication is based on the growth rate defect that results from the deletion of the ribosomal protein encoding gene *RPL20A* (*YMR242c*) in a haploid cell (doubling time of 3 h compared to 1.5 h for the wild-type (WT) strain). This gene shares 99% nucleotide sequence identity with *RPL20B* (*YOR312c*). We first showed that a supplementary copy of the *RPL20B* gene and its flanking intergenic regions (1.9 kb in total), carried by a centromeric plasmid, was indeed able to restore a normal growth phenotype (not shown). We then analyzed the appearance of spontaneous rapidly growing revertants using a Luria–Delbruck fluctuation assay (Luria and Delbruck, 1943). A total of 100 independent cultures of the slow-growing haploid parental strain deleted for *RPL20A* were propagated for approximately 80 generations by serial transfer into rich glucose liquid medium. Every eight generations, an aliquot of each culture was diluted and plated onto rich medium. After 36 h of growth on plates, spontaneous revertant strains form big colonies easily discernable from the parental small colonies. The mutation rate is estimated at approximately 10^{-10} events per cell per division (see Materials and methods). After 80 generations, 83 out of 100 cultures were entirely overtaken by revertant strains. The molecular events responsible for this growth rate reversion were first investigated by pulse-field gel electrophoresis (PFGE) karyotyping.

A majority of spontaneous normal growth revertants result from large chromosomal alterations

Among the 83 independent revertant strains isolated, we identified 48 strains showing modifications of either the size or the number of chromosomes in their karyotypes that can be classified into three classes (see Table I and Figure 1). Class I corresponds to 42 revertants showing a size increase of chromosome XV that carries *RPL20B* (Figure 1A). Class II is represented by four independent revertants that show an additional chromosome in their karyotypes (Figure 1B). Class III consists in two revertants presenting a size increase of a chromosome different from chromosome XV (Figure 1C and D). In all cases, hybridization of the karyotypes using *RPL20B* as a probe revealed either the larger or the additional chromosome, suggesting either intra- (42 cases) or interchromosomal duplications (six cases) of this gene (Figure 1).

The 35 remaining revertants showed karyotypes indistinguishable from WT using PFGE. Small duplications that would pass unnoticed on PFGE were sought by restriction mapping of the genomic DNA of these 35 strains, but no rearrangement within a 17 kb region around the *RPL20B* locus could be detected (not shown). Growth recovery in these revertants could be due to a chromosome XV disomy, but this is difficult to detect on PFGE because of the presence of the comigrating chromosome VII. To test this hypothesis, we crossed these 35 strains with a strain carrying the recessive marker *ade2* (*YOR128c*) on chromosome XV and performed tetrad analysis on the resulting diploid strains. In all, 13 strains presented a 2:2 segregation of the *ade2* phenotype

and a spore viability (88% of spore germination, s.d. 11%) identical to the control strains, consistent with a balanced number of chromosomes in their karyotypes. Suppressor mutations in these strains may correspond to point mutations in the regulatory region of *RPL20B*. The 22 other strains presented a near absence of segregation of the *ade2* marker alone (3%, s.d. 5%), fully consistent with a disomy of chromosome XV in the corresponding haploid revertant strains. Of these, 20 also show a strongly reduced viability of the meiotic products (38% of spore germination, s.d. 8%) and the remaining two strains presented moderately reduced spore viability (62 and 77%, respectively). This reduction in spore viability might result from segregation problems at meiosis caused by the presence of the additional chromosome XV in the corresponding diploids.

Altogether, the growth defect was compensated by the duplication of the remaining gene of the pair, either through segmental duplication or by aneuploidization (48 and 22 out of 83 revertant strains, respectively), or by suppressor point mutations (13 cases).

CGH analysis of revertant strains showing modified karyotypes

To determine the precise nature of the rearrangements, the genomic DNA from 26 strains belonging to class I and from the six strains belonging to classes II and III was further analyzed by comparative genomic hybridization (CGH) onto microarrays. Total DNA from each revertant strain was hybridized onto microarrays and compared to the WT strain. Genomic ratios were determined for each open reading frame (ORF) between the revertant and the WT strains and plotted as a function of their chromosomal location (Figure 1). In all strains, a series of contiguous ORFs located on the right arm of chromosome XV and encompassing *RPL20B* was found duplicated. These duplicated blocks ranged from 41 to 490 kb in size. In all, 14 different array profiles were found among the 26 strains from class I (two patterns were found 10 and four times, respectively), defining 14 different intrachromosomal duplicated blocks (Figure 1A and Table I). In addition, for the four strains carrying a supernumerary chromosome (class II), a large segment (from 133 to 655 kb) from another chromosome was also found to be duplicated (Figure 1B). In these strains, the duplicated segment from chromosome XV extends up to the right telomere, while the duplicated segment from the other chromosome includes the centromere and extends up to one telomere. In all cases, the cumulated size of the two duplicated segments identified by CGH is fully consistent with the size of the supernumerary chromosome estimated from the PFGE. Therefore, the supernumerary chromosomes result from the fusion between a duplicated part of the right arm of chromosome XV and a duplicated region from another chromosome that does not carry *RPL20B*. The remaining two strains (YKF1036 and YKF1246) belonging to class III present a size increase of chromosomes V and III, respectively (Figure 1C and D). For YKF1036, the CGH-array profile revealed a 256 kb duplication of the right arm of chromosome XV, extending up to the telomere, suggesting that this duplicated segment was fused to a complete chromosome V by a nonreciprocal translocation into one of its terminal regions. Completeness of chromosome V in the chimeric chromosome is confirmed by positive hybridizations

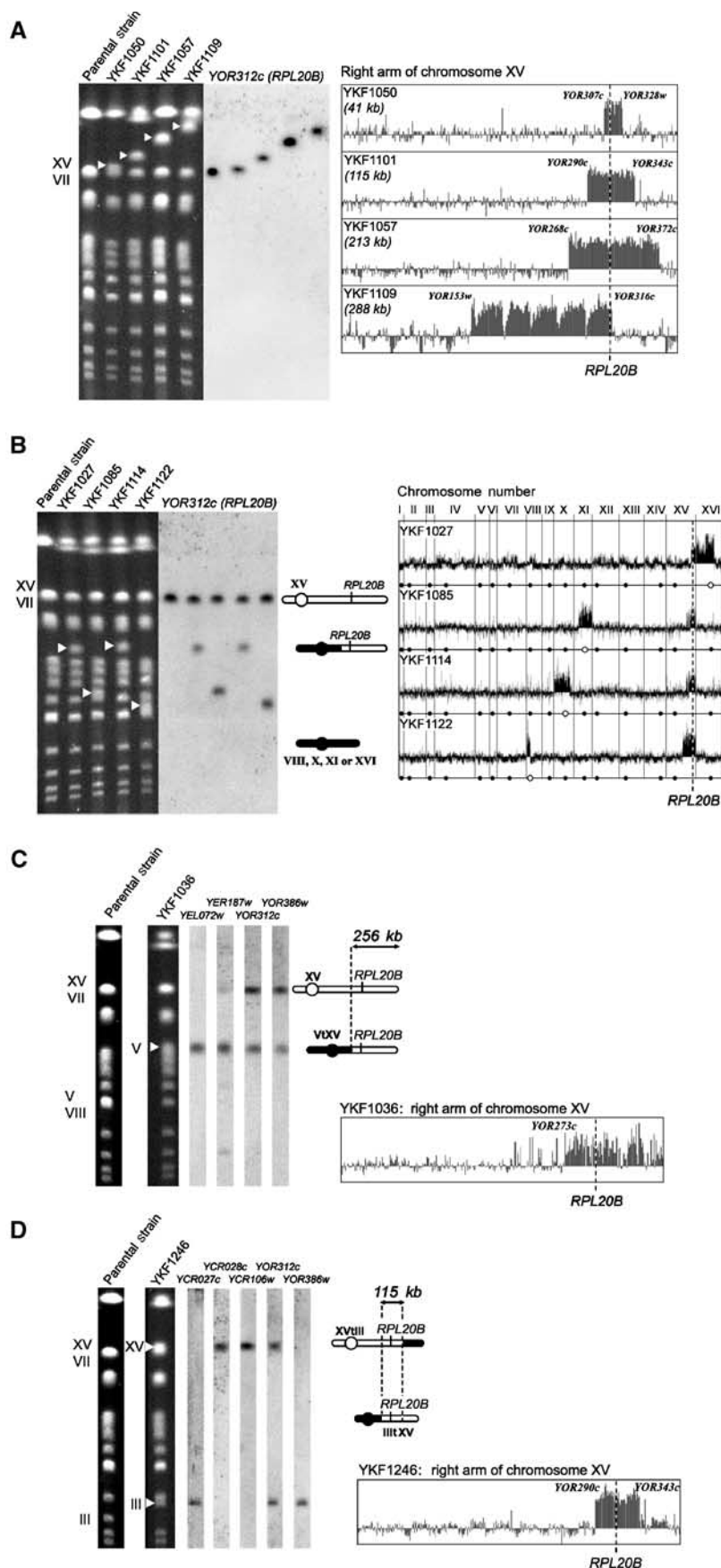


Table 1 DNA duplication events observed in revertant strains

Strain	Type (class)	Bloc size (kb)	Left border ^a	Right border ^a
YKF1109	Intra (I)	288	<i>YOR153w</i>	<i>YOR316c</i>
YKF1124	Intra (I)	263	<i>YOR220w</i>	<i>YOR362c</i>
YKF1080	Intra (I)	249	<i>YOR236w</i>	<i>YOR370c</i>
YKF1103	Intra (I)	220	<i>YOR216c</i>	<i>YOR344c</i>
YKF1057	Intra (I)	213	<i>YOR268c</i>	<i>YOR372c</i>
YKF1093	Intra (I)	213	<i>YOR268c</i>	<i>YOR372c</i>
YKF1147	Intra (I)	213	<i>YOR268c</i>	<i>YOR372c</i>
YKF1250	Intra (I)	213	<i>YOR268c</i>	<i>YOR372c</i>
YKF1110	Intra (I)	208	<i>YOR278w</i>	<i>YOR380w</i>
YKF1159	Intra (I)	183	<i>YOR269w</i>	<i>YOR357c</i>
YKF1223	Intra (I)	180	<i>YOR227w</i>	<i>YOR336w</i>
YKF1175	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1212	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1038	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1067	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1101	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1107	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1116	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1120	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1191	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1201	Intra (I)	115	<i>YOR290c</i>	<i>YOR343c</i>
YKF1072	Intra (I)	110	<i>YOR267c</i>	<i>YOR329c</i>
YKF1022	Intra (I)	102	<i>YOR272w</i>	<i>YOR328w</i>
YKF1016	Intra (I)	65	<i>YOR305w</i>	<i>YOR336w</i>
YKF1095	Intra (I)	52	<i>YOR307c</i>	<i>YOR332w</i>
YKF1050	Intra (I)	41	<i>YOR307c</i>	<i>YOR328w</i>
YKF1044	Intra (I)	~ 100 ^b		
YKF1061	Intra (I)	~ 100 ^b		
YKF1063	Intra (I)	~ 100 ^b		
YKF1065	Intra (I)	~ 100 ^b		
YKF1069	Intra (I)	~ 100 ^b		
YKF1071	Intra (I)	~ 100 ^b		
YKF1082	Intra (I)	~ 100 ^b		
YKF1112	Intra (I)	~ 100 ^b		
YKF1126	Intra (I)	~ 100 ^b		
YKF1127	Intra (I)	~ 100 ^b		
YKF1141	Intra (I)	~ 100 ^b		
YKF1043	Intra (I)	> 150 ^b		
YKF1128	Intra (I)	> 150 ^b		
YKF1132	Intra (I)	> 150 ^b		
YKF1234	Intra (I)	> 150 ^b		
YKF1286	Intra (I)	> 150 ^b		
YKF1122 ^c	Inter (II)	490 (+ 133)	<i>YHR014w</i>	<i>YOR143c</i>
YKF1114 ^c	Inter (II)	268 (+ 594)	<i>YJR090c</i>	<i>YOR267c</i>
YKF1085 ^c	Inter (II)	231 (+ 464)	<i>YKL128c</i>	<i>YOR290c</i>
YKF1027 ^c	Inter (II)	202 (+ 655)	<i>YPR044c</i>	<i>YOR305w</i>
YKF1036	Inter (III)	256	Subtelomeric V	<i>YOR273w</i>
YKF1246	Inter (III)	115	<i>YCR027c</i>	<i>YOR290c</i>

Type indicates intra- and interchromosomal events. For details on classes I–III, refer to Figure 1 and the text.

^aORFs located at the right and left borders were determined by CGH (blank: not determined).

^bSize estimated from PFGE. Other sizes are deduced from breakpoint coordinates.

^cFor strains carrying a supernumerary chromosome, the size of the duplicated segment from the other chromosome is indicated in parentheses.

Figure 1 Intra- and interchromosomal duplications of large DNA segments from the right arm of chromosome XV. Left panels show PFGE and hybridizations of the corresponding Southern blots with various ORFs whose names are indicated above each lane. Chromosomes XV and VII comigrate in the parental strain, resulting in a band with a double intensity on the gel. Relevant chromosome numbers are indicated, and the white arrowheads show the position of the modified or additional chromosomes in the karyotypes. Right panels represent CGH-array profiles with the X-axes consisting either in the ORFs from the right arm of chromosome XV ordered from *YOR001w* to *YOR394w* (A, C and D) or in all yeast ORFs ordered from the left telomere of chromosome I to the right telomere of chromosome XVI (B). The Y-axes correspond to the genomic ratios calculated between the revertant and the parental strains. Scale ranges from –0.5 to +1.5 (A, C and D) and from –1 to +2 (B). The position of *RPL20B* (*YOR312c*) is indicated by the dotted line. The last duplicated ORFs on each side of the blocks are mentioned (A, C and D). In (C) and (D), a schematic representation of the chromosomes involved in the interchromosomal duplication–translocation events is shown. (A) Representative revertant strains from class I. The sizes of the duplicated segments are indicated in parentheses and range from 41 to 288 kb. (B) Revertant strains from class II. Black and open circles on the X-axes of the array profiles symbolize single and duplicated centromeres, respectively. (C) Class III strain YKF1036. VtXV means a translocation of the right arm of chromosome XV onto chromosome V. (D) Class III strain YKF1246. IIItXV and XVtIII mean a translocation of the right arm of chromosomes XV and III onto chromosomes III and XV, respectively. The ORFs *YCR027c* and *YCR028c* flank the translocation breakpoint onto chromosome IIItXV.

of the most telomere-proximal unique ORFs (*YEL072w* and *YER187w*; Figure 1C). For YKF1246, the CGH-array profile revealed a 115 kb duplicated segment, and hybridizations of the subtelomeric probes from either chromosome III (*YCR106w*) or XV (*YOR386w*) revealed a reciprocal chromosomal arm exchange between these two chromosomes (Figure 1D). This chromosomal rearrangement corresponds to an unequal reciprocal translocation between chromosomes III and XV, leaving a 115 kb region from chromosome XV duplicated on both chromosomes. Altogether, the chromosomal rearrangements characterized in these 32 strains demonstrate that large DNA segments can be spontaneously duplicated either intra- or interchromosomally, leading to a scattering of large duplicated segments throughout the genome.

Intrachromosomal duplicated segments are repeated as direct tandems

We used DNA combing to map precisely the intrachromosomal segmental duplications within chromosome XV (Bensimon *et al*, 1994; Michalet *et al*, 1997). DNA from five recombinant cosmid clones (Tettelin *et al*, 1998), whose inserts are localized in the duplicated regions, were labeled either with biotin or digoxigenin and detected by red or green fluorochromes, respectively. Joint hybridization of these five probes onto stretched DNA from the WT strain led to an alternation of red and green signals with a defined interval length between them (Figure 2). The same hybridization onto stretched DNA from strain YKF1022, which was shown by CGH to exhibit a duplicated block of approximately 100 kb, revealed a tandemly duplicated series of additional red and green signals. The size of the duplicated segment determined by DNA combing is consistent with the chromosome XV size increase estimated from the PFGE and CGH analyses. From a similar experiment, the 65 kb duplicated segment in strain YKF1016 was also found to be repeated as a direct tandem (not shown). Thus, in the two revertants studied by DNA combing, the segmental duplications form direct tandem structures. The same is true for at least 21 of the 24 other strains with intrachromosomal duplications, as

verified by amplification and sequencing of the breakpoints (see below).

Sequences involved at the duplication breakpoints

Based on the limits of the duplicated regions determined by CGH, we amplified and sequenced the junctions of 11 intrachromosomal (corresponding to 23 strains) and four interchromosomal events. Three types of sequences were found: low-complexity DNA sequences (GTT/AAC trinucleotide repeats and polyA/T tracts), microhomology regions (from 2 to 9 nucleotides) and transposon-related sequences (Figure 3). In six cases, the junctions consisted in the fusion between two ORFs. Considering the reading frames and the coding strands, this led in three cases to the formation of a novel chimeric ORF still possessing an active promoter and composed of the 5' end of one ORF fused to the 3' end of the other (Figure 3).

We also tested whether the sequences located at both the left and right borders of the intrachromosomal duplicated blocks were modified by the duplication event (Figure 3). These regions were amplified and sequenced in five strains corresponding to the three different types of breakpoints. No changes were found compared to the corresponding WT sequences, suggesting that the external regions of the duplicated blocks are either not broken during the process or always faithfully repaired.

Generalization of the phenomenon of segmental duplications to the whole genome

Our screening procedure involved gene dosage compensation for a 1-kb-long gene located on the right arm of chromosome XV. Therefore, only the segmental duplications encompassing this gene could be observed. To test whether other chromosomal regions of the yeast genome were equally susceptible to such events, we explored deletant strains of the *Saccharomyces* Deletion Project in which the deletion of a gene was reported as resulting in a slow growth phenotype (EUROFAN, <http://mips.gsf.de/proj/yeast/CYGD/db/index.html>). Among 191 deletant strains expected to be slow growing, 134 presented no growth defect on a rich medium

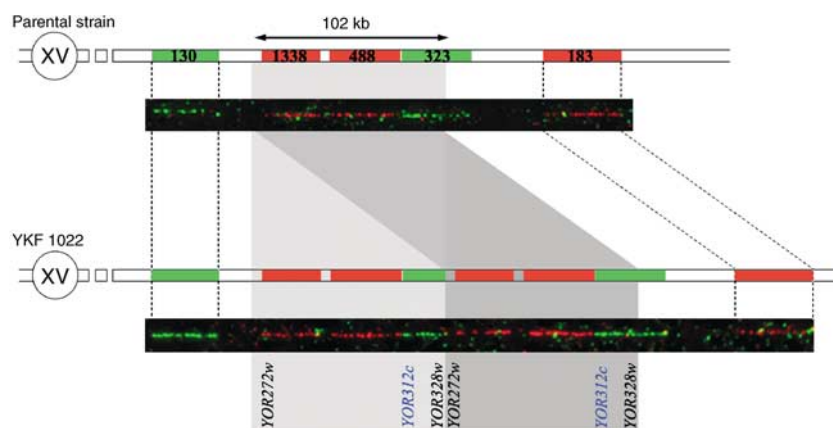


Figure 2 Direct tandem organization of an intrachromosomal duplicated segment visualized by molecular combing. Molecular combing of the genomic DNA of the parental strain (top) and of YKF1022 (bottom) performed with fluorescent probes derived from recombinant cosmids 130, 1338, 488, 323 and 183 of chromosome XV (Tettelin *et al*, 1998). Red and green colors correspond to biotin- and digoxigenin-labeled probes, respectively. The succession of the colored signals reveals that the segmental duplication of approximately 100 kb (extending from *YOR272w* up to *YOR328w* as deduced from CGH) is oriented as a direct tandem. These patterns were analyzed on nine and 12 different DNA molecules for the WT and the YKF1022 strains, respectively.

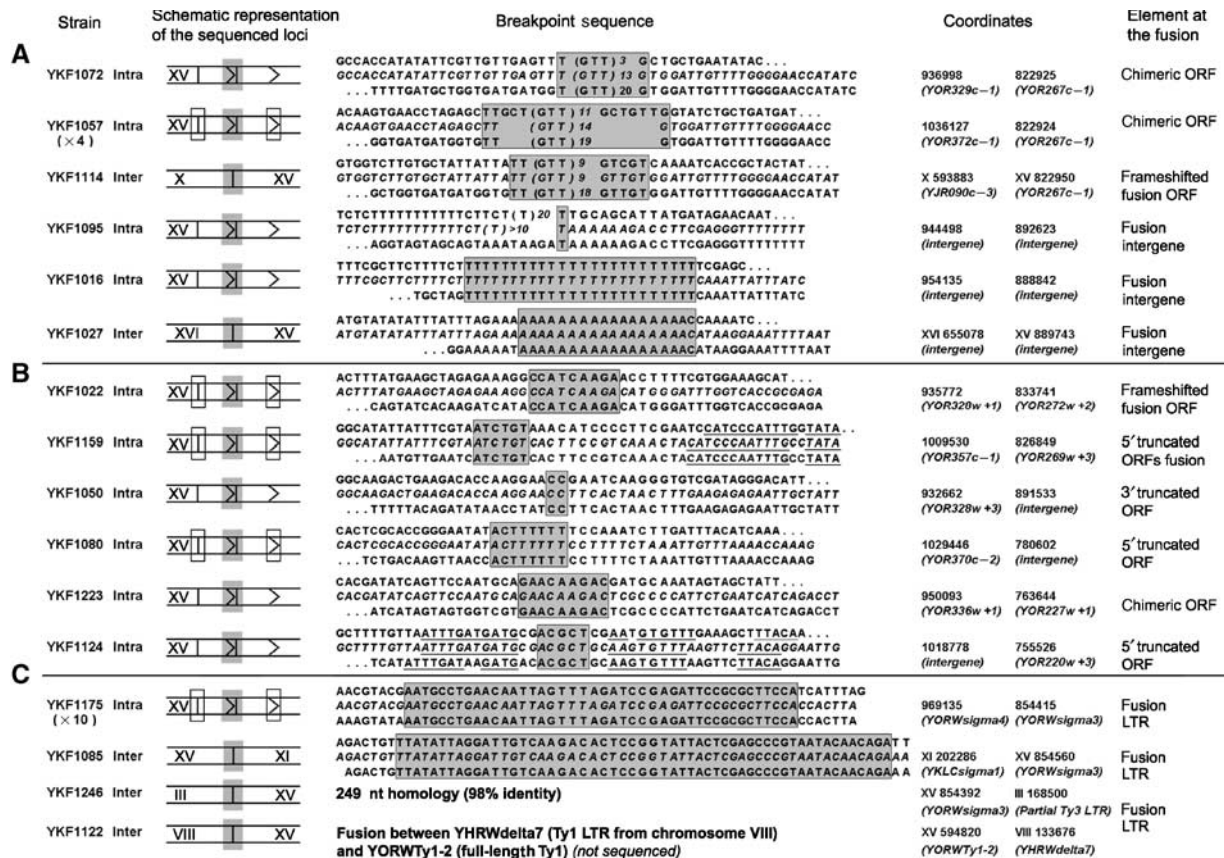


Figure 3 Segmental duplication breakpoints: (A) microsatellites junctions; (B) microhomology junctions; and (C) transposon-related junctions. Gray boxes represent the localization of the junctions. For intrachromosomal events (Intra), the top and bottom sequences are the flanking telomere- and centromere-proximal sequences of the junctions, respectively. For interchromosomal duplications (Inter), top sequences correspond to the centromere-containing chromosomes whereas bottom sequences represent the right arm of chromosome XV. Sequences of the junctions are italicized in the middle lines. Microhomology regions flanking the junctions are underlined. The genetic elements involved in the junctions are mentioned below the coordinates, and the frame is specified when the breakpoint occurred within an ORF. Empty boxes on the schematic representations represent the sequenced left and right borders of some intrachromosomal duplicated blocks. In these sequences, no change was found compared to the corresponding WT sequences.

and 27 of them showed large karyotypic modifications (see Table II). A total of 20 strains were found to be disomic for one chromosome, and in 17 such cases this chromosome carries a close homolog to the deleted gene. In the seven other strains, a size increase of one chromosome was observed. This chromosome carries a sequence closely related to the deleted gene in three out of the seven cases, suggesting intrachromosomal duplications of a DNA segment. In the four remaining strains, the larger chromosome was different from the one that carries the closest paralogous gene, suggesting that interchromosomal duplication events might have occurred. These results confirm previous findings on the high level of aneuploidy within deletant yeast strain collections (Hughes *et al*, 2000). They also show that the mechanism of segmental duplication described in this work for the chromosomal region encompassing *RPL20B* can be generalized to other chromosomes and, possibly, to the whole genome.

Discussion

Using a single gene dosage compensation system, we have been able to identify several large segmental duplication events all encompassing the reporter gene, but extending

much beyond it. It should be stressed that although duplication of this gene confers a significant growth advantage, the mutational events studied here are spontaneous ones, and not induced. Most events correspond to intrachromosomal direct tandem duplications of large segments from chromosome XV (42/48). In the six remaining cases, interchromosomal duplication events led to the formation of chimeric chromosomes.

It is very significant for molecular evolution that, although our selection system acted on a single gene, large chromosomal segments encompassing many genes were always duplicated concomitantly with it (up to 252 ORFs from chromosome XV and 327 ORFs from other chromosomes). Such pools of duplicated genes may represent important reservoirs of coding sequences, possibly allowing the emergence of new functions through modification of their coding and/or the regulatory sequences (Lynch, 2002). Another possibility for the emergence of a new function is the formation of fusion genes by domain accretion at the junction of the duplicated blocks (Eichler, 2001; Bailey *et al*, 2002). We characterized chimeric ORFs at the junctions of six of the duplication events. Similarly, in primate species, Courseaux *et al* (2003) showed that euchromatic segmental duplications

Table II Karyotype modifications in deleted strains from the *Saccharomyces* Deletion Project library

Strain number ^a	Deleted ORF	Disomy	Segmental duplication	Possible explanation for growth recovery
16679 (α)	YBR191w	XVI		Duplication of YPL079w
16961 (α)	YBR189w	XVI		Duplication of YPL081w
6961 (a)	YBR189w	XVI		Duplication of YPL081w
15813 (α)	YCR094w	XIV		Duplication of either YNL323w or YNR048w
5813 (a)	YCR094w	XIV		Duplication of either YNL323w or YNR048w
16475 (α)	YOR096w	XIV		Duplication of YNL096c
17002 (α)	YJL136c	XI		Duplication of YKR057w
7002 (a)	YJL136c	XI		Duplication of YKR057w
16595 (α)	YHL011c	V or VIII		Duplication of YER099c
17224 (α)	YNL079c	IX		Duplication of YIL138c
17376 (α)	YNL069c	IX		Duplication of YIL133c
37376 (a/α)	YNL069c	IX		Duplication of YIL133c
7224 (a)	YNL079c	IX		Duplication of YIL138c
7374 (a)	YNL055c	IX		Duplication of YIL114c
7376 (a)	YNL069c	IX		Duplication of YIL133c
17220 (α)	YNL073w	IV and III		Duplication of YDR037w
16115 (α)	YER117w	II		Duplication of YBL087c
6589 (a)	YGR188c	III		Segregation problem of YGR188c (BUB1)
5753 (a)	YCR024c-a	XI		Unknown
5756 (a)	YCR027c	XI		Unknown
6847 (a)	YJR049c		On XVI	Intrachromosomal duplication encompassing YPL188c
35278 (a/α)	YLR369w		On X	Intrachromosomal duplication encompassing YJR045c
26955 (a/α)	YML085c		On XIII	Intrachromosomal duplication encompassing YML124c
5278 (a)	YLR369w		On XVI	Interchromosomal duplication encompassing YJR045c
5731 (a)	YBR283c		On XVI	Interchromosomal duplication encompassing YLR378c
16897 (α)	YJR077c		On I	Interchromosomal duplication encompassing YER053c
16990 (α)	YGR092w		On V or VIII	Interchromosomal duplication encompassing YPR11w

^aMating types are indicated in parentheses.

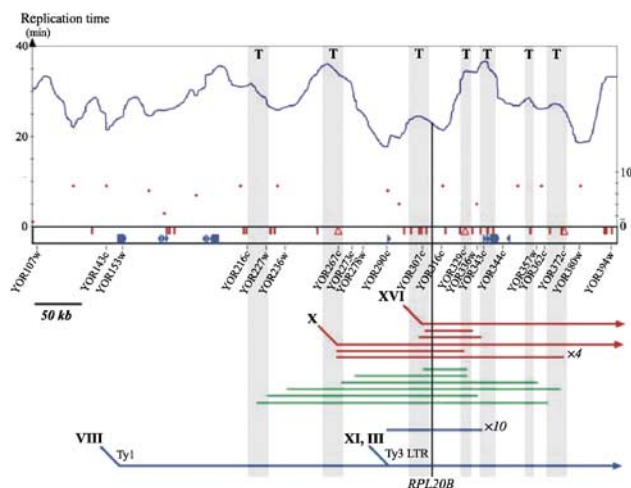


Figure 4 Mapping of duplicated segment end points relative to both the replication timing curve and the genetic elements of the region. (Top) Replication time data are from Raghuraman *et al* (2001). The X-axis represents a part of the right arm of chromosome XV, extending from YOR107w up to the telomere. Red diamonds symbolize replication origins with their prediction confidence level, ranging from 1 (poor confidence value) to 9 (strong confidence value), indicated in the right Y-axis. T's within shaded columns symbolize the major replication termination sites flanking RPL20B. Blue arrowheads are for LTRs, and oriented rectangles stand for full-length Ty elements. Red bars correspond to runs of A/T longer than 14 base pairs. Red void triangles represent GTT/AAC trinucleotide repeats longer than three units. (Bottom) Duplicated segments are symbolized by the colored lines according to the type of sequences found at their junctions: red for microsatellites, green for microhomology and blue for transposable elements. Arrowheads indicate that the segment extends to the telomere. Interchromosomal events are indicated by upper diagonals with chromosome names. The vertical black line indicates the position of YOR312c (RPL20B). Two segmental duplication events were respectively found four and 10 times.

had created mosaic genes at their junctions. Thus, besides generating a large pool of genes onto which evolution can act, large segmental duplications also provide opportunities for the emergence of new functions through the formation of mosaic genes at their boundaries. Copy number increase as well as chimeric ORF formation is reminiscent of the genomic disorders associated with some genetic diseases and malignant transformations of higher eucaryotic cells (Lupski, 1998; Abdel-Rahman *et al*, 2001).

The molecular mechanisms responsible for the formation of such large segmental duplications are obviously of major interest to determine. If unequal crossing over (UCO) is generally invoked to explain the formation of duplicated regions, it is unlikely here that this mechanism has played a role as no large sequence similarity was found at the junction of 60% of the duplicated blocks (microhomology and microsatellite sequences). Even if UCO could be invoked to explain the duplications between the Ty long terminal repeats (LTRs), we must also underline that many other repeated sequences (LTRs and full-length transposons) that are present in the region encompassed by the duplications were never used as substrate for recombination. Rather, the replication map of the right arm of chromosome XV (Raghuraman *et al*, 2001) shows that all the termination sites flanking the RPL20B locus were involved at least once in our segmental duplication breakpoint (Figure 4). In addition, all of the duplicated segments, apart from the three Ty-mediated interchromosomal events, presented at least one end located in the vicinity of a termination fork site on chromosome XV (Figure 4). This is consistent with recent studies in yeast showing that termination sites, microsatellites and transposon-rich regions attenuate the progression of the fork, causing substantial replication stalling and thereby

offering a potential source of genome instability (Cha and Kleckner, 2002; Pelletier *et al*, 2003). In a normal cell, numerous and redundant pathways are aimed at maintaining genome stability (Kolodner *et al*, 2002). For example, S-phase checkpoints suppress a large diversity of gross chromosomal rearrangements that result from replication accidents (Myung *et al*, 2001). It is then tempting to propose that segmental duplications would occur by a two-step mechanism. First, a replication fork pauses and collapses generating a chromosome breakage. Second, the double-strand break can be processed into a new replication fork either intra- or intermolecularly by a break-induced replication-like mechanism that does not necessarily need a long sequence homology (Malkova *et al*, 1996; Morrow *et al*, 1997; Bosco and Haber, 1998; Ira and Haber, 2002). Alternatively, direct tandem segmental duplication could result from the cleavage of the two strands of the same polarity from two replication forks flanking *RPL20B* followed by a simple nonhomologous end joining event between two DNA ends.

The size of our segmental duplications is important to consider with regard to the mechanisms of genome evolution. The smallest segment observed here extends over 41 kb and the largest over 490 kb. This size distribution (mean = 172 kb, σ = 89 kb) is larger than the size distribution of the ancestral duplication blocks recognized in the *S. cerevisiae* genome (mean = 58 kb, σ = 46 kb; Wolfe and Shields, 1997). However, the presently recognized size of the duplicated blocks in the yeast genome is necessarily smaller than their original extent, due to the massive gene loss that has been occurring within these regions since their formation (Fischer *et al*, 2001). In addition, the orientation of the duplicated segments that we observe is always conserved relative to the centromeres, as it is mainly the case for the ancestral blocks of duplication. Our experimental results suggest that, as proposed earlier from a comparative genomic analysis (Llorente *et al*, 2000a), the present structure of the *S. cerevisiae* genome may have been reshaped by the successive accumulation of segmental duplications over evolutionary times.

Finally, the significance of our results for the general understanding of the duplication process during evolution may be questioned because we imposed a strong selection for duplication events encompassing a functional ribosomal protein gene. However, our data (Table II) and those from Hughes *et al* (2000) indicate that similar large duplication events are found in other parts of the genome using a variety of other deletion mutants. It is therefore possible that any part of the WT genome may be subject to such spontaneous duplication events, which may eventually be fixed in a natural population by selection or isolation. The quantitative importance of such events in the evolution of natural populations obviously remains to be determined. Ongoing sequencing programs of yeast genomes may rapidly solve this question.

Materials and methods

Yeast strains and plasmids

All strains are derivatives of *S. cerevisiae* BY4743 (Winzler *et al*, 1999). Heterozygous diploid strains YKF118, YKF119 and YKF120 were constructed by PCR-based gene replacement of one *YMR242c*

(*RPL20A*) allele by the *KanMX4* resistance gene (Wach *et al*, 1994). Slow-growing, geneticin-resistant meiotic products of YKF118, YKF119 and YKF120 were used as parental strains. A detailed list of the 191 deleted strains tested from the *Saccharomyces* Deletion Project is available on request.

To test whether an additional copy of *RPL20B* was able to restore a normal growth rate to the parental strains, the *RPL20B* gene and its flanking intergenic regions (1.9 kb in total) were cloned into the centromeric pRS416 plasmid (Stratagene).

The 35 revertant strains carrying no large duplications were crossed to Y55 strains (either Y2366 (*Mata, ade2*) or Y2367 (*Matalpha, ade2*)). Control strains were constructed crossing BY4741 with Y2367 and BY4742 with Y2367. The diploids were sporulated and 8–18 tetrads were dissected for each strain.

Growth selection assay and mutation rate calculation

A total of 100 independent cultures, each issued from one independent subclone from our slow growth parental haploid *RPL20A::KanMX* strains, were inoculated with 2×10^7 cells in YPD (glucose rich) medium, at 30°C, under agitation. Approximately 5×10^6 cells from each culture were transferred every eight generations in fresh YPD medium. This process was repeated 10 times for a total of 80 generations. During each of these successive culture cycles, exponentially growing cells were plated on YPD, at a density of 2×10^2 to 2×10^3 cells per plate. The first normal growing revertant colonies, showing a much bigger size than their ancestor, appeared on plates after 120 h (20 generations) in five cultures and kept appearing in other cultures until the experiment was stopped (at this time, 83 cultures were overtaken by normal growth revertants). For the first revertant cells to be detectable on the plates, the proportion of revertants in the culture must be approximately 0.5%. Based on the growth rate difference between parental and revertant cells (0.33 and 0.66 doubling/h, respectively) and on the total number of cells transferred through each propagation step (5×10^6 cells), we estimated that the initial reversion event must have occurred 3–4 generations after the initial inoculation step to reach this proportion. As only five out of the 100 cultures presented revertants on plates after 120 h of cultivation, the mutation rate was calculated from a Luria–Delbruck fluctuation test considering the probability of having no mutations in 95 out of 100 cultures ($f = 0.95$) after 3–4 generations starting from an initial number of cells $n_0 = 2 \times 10^7$. This corresponds to a total number of mitotic divisions of $n = 1.6 \times 10^8$ to 3.2×10^8 ($2 \times 10^7 \times 2^3$ to $2 \times 10^7 \times 2^4$). The mutation rate is given by the probability $p = 1 - e^{lnf/(n-n_0)}$ corresponding to 1.7×10^{-10} to 3.7×10^{-10} reversion events per cell per division.

In each culture, at least two (and up to six) independent revertant colonies were picked and subcloned for molecular analysis. In total, 287 normal growth revertants were studied originating from at least 83 independent mutational events. In all cases, all the revertant colonies isolated from the same culture presented the same karyotype and, therefore, were considered as descending from the same mutational event.

Pulsed-field gel electrophoresis

PFGE karyotypes were established in a Rotaphor R23 (Biometra) or in a CHEF Mapper XA (Bio-Rad) tank and transferred onto Hybond N+ membranes (Amersham) according to the manufacturer's recommendations.

Comparative genomic hybridizations and DNA combing

Total genomic DNA of mutants and parental strains was prepared with the Qiagen genomic TIP20 and hybridized against yeast whole-genome arrays from Eurogentec or Affymetrix. For Eurogentec arrays, 0.5–1 µg of genomic DNA was labeled by random priming with fluorescent Cy3- (parental strain) and Cy5-dUTP (revertants) as previously described (Hughes *et al*, 2000). Genomic ratio for each ORF was defined as the ratio between normalized spot intensities of the revertant and parental strains, from which the mean of all spot intensities ratios was subtracted. For Affymetrix arrays (YG-S98), labeling, hybridization and detection steps were performed at the Affymetrix Platform from the G  n  pole Alsace-Lorraine (IGBMC Illkirch, France). Arrays were analyzed and genomic ratios were calculated by the Affymetrix GeneChip software.

The DNA combing procedure is detailed in Conti *et al* (2001). Briefly, genomic DNA of strains YKF1016 and YKF1022 was purified in agarose plugs, and stained with YOYO-1 (Molecular Probes). Plugs were melted at 65°C and digested with β -agarase. The DNA solution, in 50 mM MES (pH 5.5), was poured in a Teflon reservoir. Silanized coverslips were dipped into the DNA solution for 5 min and pulled out at a constant speed (0.3 mm/s). *In situ* hybridization of fixed DNA was performed using cosmids from the right arm of chromosome XV as probes. For both strains, cosmids 130, 1338, 488, 323 and 183 (Tettelin *et al*, 1998) were labeled either with biotin-dUTP (1338, 488 and 183) or dig-dUTP (130 and 323). Signals were amplified by several consecutive layers of fluorophore-conjugated antibodies and finally detected with Texas red (for biotin probes) or FITC (for digoxigenin probes). Slides were scanned using an epifluorescence microscope connected to a CCD camera.

References

- Abdel-Rahman WM, Katsura K, Rens W, Gorman PA, Sheer D, Bicknell D, Bodmer WF, Arends MJ, Wyllie AH, Edwards PA (2001) Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement. *Proc Natl Acad Sci USA* **98**: 2538–2543
- Bach ML, Roelants F, De Montigny J, Huang M, Potier S, Souciet JL (1995) Recovery of gene function by gene duplication in *Saccharomyces cerevisiae*. *Yeast* **11**: 169–177
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. *Science* **297**: 1003–1007
- Bainbridge BW, Roper JA (1966) Observations on the effects of a chromosome duplication in *Aspergillus nidulans*. *J Gen Microbiol* **42**: 417–424
- Bensimon A, Simon A, Chiffaudel A, Croquette V, Heslot F, Bensimon D (1994) Alignment and sensitive detection of DNA by a moving interface. *Science* **265**: 2096–2098
- Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res* **13**: 137–144
- Bosco G, Haber JE (1998) Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* **150**: 1037–1047
- Cha RS, Kleckner N (2002) ATR homolog *Mec1* promotes fork progression, thus averting breaks in replication slow zones. *Science* **297**: 602–606
- Conti C, Caburet S, Schurra C, Bensimon A (2001) Molecular combing. In *Current Protocols in Cytometry*, pp 8.10.11–18.10.23. New York: John Wiley & Sons, Inc.
- Corseaux A, Richard F, Grosgeorge J, Ortola C, Viale A, Turc-Carel C, Dutrillaux B, Gaudray P, Nahon JL (2003) Segmental duplications in euchromatic regions of human chromosome 5: a source of evolutionary instability and transcriptional innovation. *Genome Res* **13**: 369–381
- Daud F, Ortori GS, Roper JA (1985) Spontaneous IR duplications generated at mitosis in *Aspergillus nidulans*: further evidence of a preferential site of transposed attachment. *Genetics* **110**: 229–245
- Dorsey MJ, Hoeh P, Paquin CE (1993) Phenotypic identification of amplifications of the *ADH4* and *CUP1* genes of *Saccharomyces cerevisiae*. *Curr Genet* **23**: 392–396
- Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **99**: 16144–16149
- Eichler EE (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet* **17**: 661–669
- Fischer G, Neuveglise C, Durrens P, Gaillardin C, Dujon B (2001) Evolution of gene order in the genomes of two related yeast species. *Genome Res* **11**: 2009–2019
- Friedman R, Hughes AL (2001) Pattern and timing of gene duplication in animal genomes. *Genome Res* **11**: 1842–1847
- Friedman R, Hughes AL (2003) The temporal distribution of gene duplication events in a set of highly conserved human gene families. *Mol Biol Evol* **20**: 154–161
- Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, Friend SH, Marton MJ

Acknowledgements

We are very grateful to A Bensimon for his help with the molecular combing technology. We thank C Thibault from the 'Génopole Alsace-Lorraine' for the Affymetrix array hybridizations and JY Coppée and O Sismeiro from the 'Génopole Institut Pasteur' for their valuable help with PCR product-based arrays. We thank Ed J Louis and Gianni Litti for providing us with the Y55 strains. We also thank GF Richard for a critical reading of the manuscript, and our colleagues from the Génolevures network and from the Unité de Génétique Moléculaire des Levures for fruitful discussions. This work was supported in part by the CNRS (GDR 2354 'Génolevures II'). Purchase of the Affymetrix arrays was supported by the French 'Ministère de la Recherche'. RK is the recipient of a CNRS BDI fellowship. BD is a member of the Institut Universitaire de France.

- (2000) Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet* **25**: 333–337
- Ira G, Haber JE (2002) Characterization of *RAD51*-independent break-induced replication that acts preferentially with short homologous sequences. *Mol Cell Biol* **22**: 6384–6392
- Kolodner RD, Putnam CD, Myung K (2002) Maintenance of genome stability in *Saccharomyces cerevisiae*. *Science* **297**: 552–557
- Llorente B, Durrens P, Malpertuy A, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, Dujon B, de Montigny J, Lepingle A, Neuveglise C, Ozier-Kalogeropoulou O, Potier S, Saurin W, Tekala F, Toffano-Nioche C, Wesolowski-Louvel M, Wincker P, Weissenbach J, Souciet J, Gaillardin C (2000a) Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. *FEBS Lett* **487**: 122–133
- Llorente B, Malpertuy A, Neuveglise C, de Montigny J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, Durrens P, Gaillardin C, Lepingle A, Ozier-Kalogeropoulou O, Potier S, Saurin W, Tekala F, Toffano-Nioche C, Wesolowski-Louvel M, Wincker P, Weissenbach J, Souciet J, Dujon B (2000b) Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. *FEBS Lett* **487**: 101–112
- Locke DP, Segreaves R, Carbone L, Archidiacono N, Albertson DG, Pinkel D, Eichler EE (2003) Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res* **13**: 347–357
- Lupski JR (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* **14**: 417–422
- Luria S, Delbruck M (1943) Mutation of bacteria from virus sensitive to virus resistant. *Genetics* **28**: 491–511
- Lynch M (2002) Genomics. Gene duplication and evolution. *Science* **297**: 945–947
- Malkova A, Ivanov EL, Haber JE (1996) Double-strand break repair in the absence of *RAD51* in yeast: a possible role for break-induced DNA replication. *Proc Natl Acad Sci USA* **93**: 7131–7136
- Michael X, Ekong R, Fougereousse F, Rousseaux S, Schurra C, Hornigold N, van Slegtenhorst M, Wolfe J, Povey S, Beckmann JS, Bensimon A (1997) Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science* **277**: 1518–1523
- Moore IK, Martin MP, Dorsey MJ, Paquin CE (2000) Formation of circular amplifications in *Saccharomyces cerevisiae* by a breakage-fusion-bridge mechanism. *Environ Mol Mutagen* **36**: 113–120
- Morrow DM, Connelly C, Hieter P (1997) 'Break copy' duplication: a model for chromosome fragment formation in *Saccharomyces cerevisiae*. *Genetics* **147**: 371–382
- Myung K, Datta A, Kolodner RD (2001) Suppression of spontaneous chromosomal rearrangements by S phase checkpoint functions in *Saccharomyces cerevisiae*. *Cell* **104**: 397–408
- Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM (2003) Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol* **23**: 1349–1357
- Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet* **3**: 827–837

- Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL (2001) Replication dynamics of the yeast genome. *Science* **294**: 115–121
- Robinson-Rechavi M, Marchand O, Escriva H, Laudet V (2001) An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. *Curr Biol* **11**: R458–R459
- Roelants F, Potier S, Souciet JL, de Montigny J (1995) Reactivation of the ATCase domain of the *URA2* gene complex: a positive selection method for Ty insertions and chromosomal rearrangements in *Saccharomyces cerevisiae*. *Mol Gen Genet* **246**: 767–773
- Samonte RV, Eichler EE (2002) Segmental duplications and the evolution of the primate genome. *Nat Rev Genet* **3**: 65–72
- Sexton CE, Roper JA (1984) Spontaneous duplications and transpositions of a large chromosome segment in *Aspergillus nidulans*. *J Gen Microbiol* **130** (Part 3): 583–595
- Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van De Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **99**: 13627–13632
- Spector LM, Fogel S (1992) Mitotic hyperploidy for chromosomes VIII and III in *Saccharomyces cerevisiae*. *Curr Genet* **21**: 309–318
- Tettelin H, Thierry A, Goffeau A, Dujon B (1998) Physical mapping of chromosomes VII and XV of *Saccharomyces cerevisiae* at 3.5 kb average resolution to allow their complete sequencing. *Yeast* **14**: 601–616
- Thomas JW, Schueler MG, Summers TJ, Blakesley RW, McDowell JC, Thomas PJ, Idol JR, Maduro VV, Lee-Lin SQ, Touchman JW, Bouffard GG, Beckstrom-Sternberg SM, Program NC, Green ED (2003) Pericentromeric duplications in the laboratory mouse. *Genome Res* **13**: 55–63
- Wach A, Brachat A, Pohlmann R, Philippsen P (1994) New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**: 1793–1808
- Whittaker SG, Rockmill BM, Blechl AE, Maloney DH, Resnick MA, Fogel S (1988) The detection of mitotic and meiotic aneuploidy in yeast using a gene dosage selection system. *Mol Gen Genet* **215**: 10–18
- Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentale E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Davis RW *et al* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**: 901–906
- Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708–713
- Wong S, Butler G, Wolfe KH (2002) Gene order evolution and paleopolyploidy in hemiascomycete yeasts. *Proc Natl Acad Sci USA* **99**: 9272–9277